View Reviews

Paper ID
1112

Paper Title
PeR-ViS: Person Retrieval in Video Surveillance using Semantic Description

Track Name
Second Round Submission

Reviewer #1

Questions

1. APPLICATIONS OR ALGORITHMS PAPER? Ideally the authors indicated which track in the CMT version of their title. We'll send an email shortly if not (you probably can't see it otherwise). Note for your evaluation that Applications papers will generally have less in the way of theoretical foundations/approach, more in experimental evaluation. Algorithms papers will be the opposite, more in the theory/approach, perhaps slightly less experiments.
Algorithms Track

2. PAPER SUMMARY What is the paper about? Please, be concise (2 to 3 sentences).
The paper presents a method for person retrieval from surveillance style videos. The main contribution of the paper is the use of cascade filtering to reduce search complexity. They compare results against some methods against [14] dataset.

3. PAPER STRENGTHS Please discuss, justifying your comments with the appropriate level of details, the strengths of the paper (i.e. novelty, theoretical approach and/or technical correctness, adequate evaluation, clarity, etc).
The paper in most parts is complete, that is, they describe the problem, the proposed solution (other than blurry diagrams), methodology, analysis.

4. PAPER WEAKNESSES Please discuss, justifying your comments with the appropriate level of details, the weaknesses of the paper (i.e. lack of novelty – given references to prior work-, lack of novelty, technical errors, or/and insufficient evaluation, etc). If you think there is an error in the paper, please explain why it is an error.
- The diagrams texts is non readable, please fix it.
- Please get the paper reviewed for an English Grammar check
- Table 2 lists Yaguchi et al. [12] three times, please explain why there.
- The ablation does not include descriptors ablation (e.g. only torso, only height).

5. FINAL RECOMMENDATION Please use "Borderline" sparingly.
Borderline

6. JUSTIFICATION Justify your final recommendation based on the strengths and weaknesses. Please be considerate to the authors and provide constructive feedback.

- The justification for soft features chosen is missing. While the authors argue that using Height, Torso, etc. were effective empirically, why would this method hold better than those in [35] and Beyond Part Models: Person Retrieval with Refined Part Pooling Sun et al.?
- While SOTA discusses several important papers Person search, and Retrieval.
- The authors choose to compare against a v small subset of SOTA. Please justify why these methods were chosen? Similarly, why does SoftBioSearch dataset make most sense?
- Usage of Mask R-CNN has already been shown in [12], thus is not a novel contribution. Please rewrite the contributions as only last one [Line 230] is novel.
- Significance of Fig 5 is unclear.
- If possible include video results

Reviewer #2

Questions

1. APPLICATIONS OR ALGORITHMS PAPER? Ideally the authors indicated which track in the CMT version of their title. We'll send an email shortly if not (you probably can't see it otherwise). Note for your evaluation that Applications papers will generally have less in the way of theoretical foundations/approach, more in experimental evaluation. Algorithms papers will be the opposite, more in the theory/approach, perhaps slightly less experiments.
Algorithms Track

2. PAPER SUMMARY What is the paper about? Please, be concise (2 to 3 sentences).
Authors propose a method for semantic-based retrieval of persons in videos. To this end, they support a number of semantic attributes and allows for generalizing person re-identification to identification with semantic attributes.

3. PAPER STRENGTHS Please discuss, justifying your comments with the appropriate level of details, the strengths of the paper (i.e. novelty, theoretical approach and/or technical correctness, adequate evaluation, clarity, etc).
- The paper tackles an interesting problem, while authors seem to provide enough justification to demonstrate that the proposed method indeed works.
- Several pre-processing steps are used to ensure that persons will be correctly processed.
- Material is provided to ensure the reproducibility of the research.

4. PAPER WEAKNESSES Please discuss, justifying your comments with the appropriate level of details, the weaknesses of the paper (i.e. lack of novelty – given references to prior work-, lack of novelty, technical errors, or/and insufficient evaluation, etc). If you think there is an error in the paper, please explain why it is an error.
- The semantic attributes are limited to "torso color, type and pattern", to leg color and pattern, as well as gender. This methods does not seem to support more common attributes, like age, body type-weight, hair color, etc. which would be also a way I would expect a human would describe a person that would need to be found

5. FINAL RECOMMENDATION Please use "Borderline" sparingly.
Weak Accept

6. JUSTIFICATION Justify your final recommendation based on the strengths and weaknesses. Please be considerate to the authors and provide constructive feedback.
I think this is an overall good paper, which some weaknesses, e.g., a limited set of attributes is support, no textual description are supported, no end-to-end learning for most of its parts. So, my recommendation is to accept this paper, since it also motivates several interesting future research directions on the topics.

Reviewer #3

Questions

1. APPLICATIONS OR ALGORITHMS PAPER? Ideally the authors indicated which track in the CMT version of their title. We'll send an email shortly if not (you probably can't see it otherwise). Note for your evaluation that Applications papers will generally have less in the way of theoretical foundations/approach, more in experimental evaluation. Algorithms papers will be the opposite, more in the theory/approach, perhaps slightly less experiments.
Applications Track

2. PAPER SUMMARY What is the paper about? Please, be concise (2 to 3 sentences).
This paper focuses on retrieving a specific person with the query of semantic description in video surveillance. The motivation of this paper is that we often can find an image of an interested person as a query. The authors proposed a deep learning-based cascade filtering approach. They first used Mask R-CNN to segment persons in the video and then utilized the DenseNet-161 to do classification.

3. PAPER STRENGTHS Please discuss, justifying your comments with the appropriate level of details, the strengths of the paper (i.e. novelty, theoretical approach and/or technical correctness, adequate evaluation, clarity, etc).
1) The method is well-illustrated. It is simple and easy to follow.
2) Figure 7 well demonstrates the advantage of the proposed method step by step.

4. PAPER WEAKNESSES Please discuss, justifying your comments with the appropriate level of details, the weaknesses of the paper (i.e. lack of novelty – given references to prior work-, lack of novelty, technical errors, or/and insufficient evaluation, etc). If you think there is an error in the paper, please explain why it is an error.
1) The methods used in this paper seem a little common. No new methods were designed.
2) I would like to know if the filer processes have their order, namely, we can filter height, torso, leg, and gender one by one, and can not change the order.
3) The height filter seems a little unreasonable, since a person may have different heights in different cameras, due to the distance to the camera and capturing angle.
4) The authors did not illustrate how to train the classification models and what dataset is used. The compared baseline algorithm is also not clear in the paper.
5) From Table 1, we can see that the validation accuracies of different attributes are not so good. Please try to train the models well with other backbones, such as Resnet-50.

5. FINAL RECOMMENDATION Please use "Borderline" sparingly.
Borderline

6. JUSTIFICATION Justify your final recommendation based on the strengths and weaknesses. Please be considerate to the authors and provide constructive feedback.
An effective application, but the method is somehow very common as research work and not so novel.

Reviewer #4

Questions

1. APPLICATIONS OR ALGORITHMS PAPER? Ideally the authors indicated which track in the CMT version of their title. We'll send an email shortly if not (you probably can't see it otherwise). Note for your evaluation that Applications papers will generally have less in the way of theoretical foundations/approach, more in experimental evaluation. Algorithms papers will be the opposite, more in the theory/approach, perhaps slightly less experiments.
Applications Track

2. PAPER SUMMARY What is the paper about? Please, be concise (2 to 3 sentences).
This paper describes a method of person search by soft biometrics. The proposed system first extracts person instances by Mask R-CNN and then extracts color, type, pattern from torso and leg modules, gender with DenseNet, while it extracts the height by camera calibration. Experiments with SoftBioSearch shows the effectiveness of the proposed method.

3. PAPER STRENGTHS Please discuss, justifying your comments with the appropriate level of details, the strengths of the paper (i.e. novelty, theoretical approach and/or technical correctness, adequate evaluation, clarity, etc).
- The proposed system outperforms some benchmarks for SoftBioSearch data set.

4. PAPER WEAKNESSES Please discuss, justifying your comments with the appropriate level of details, the weaknesses of the paper (i.e. lack of novelty – given references to prior work-, lack of novelty, technical errors, or/and insufficient evaluation, etc). If you think there is an error in the paper, please explain why it is an error.
- Important references are missing.

- The proposed system is not well differentiated from the previous work.

- Experimental validation is insufficient.

- There are many editorial errors.

More details on the above-mentioned concerns are as follows.

1. Important references are missing.

The authors mainly differentiate the proposed system from methods of image query-based person search (e.g., person re-identification) in the introduction. The reviewer is afraid that the authors overlooked literature on attribute/language query-based person search with zero-short learning (i.e., zero-short person search) including but not limited to the followings.

[39] Q. Dong, X. Zhu and S. Gong, "Person Search by Text Attribute Query As Zero-Shot Learning," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 3651-3660, doi: 10.1109/ICCV.2019.00375.

Moreover, technical aspects of what the authors want to do in this paper are closely relevant with zero-shot learning, so the reviewer would suggest the authors broadly survey the papers on zero-shot learning.

2. The proposed system is not well differentiated from the previous work.

The authors cite a couple of papers on person search with attribute or languages in the introduction section (e.g., [2, 3, 5]). Despite of the close relevance, the authors do not sufficiently differentiate the proposed method from them. The reviewer would suggest the authors to clearly state the drawbacks/limitations of the existing approaches and then how the proposed method overcome the drawbacks/limitations, in order to well justify the proposed method.

3. Experimental validation is insufficient.

The authors conducted experiments with a single data set (i.e., SoftBioSearch). Although the reviewer understands that SoftBioSearch is one of suitable data sets for person search by attribute, the variety of benchmarks for this data set seems to be limited. So, the reviewer would suggest the authors to validate the proposed method over other data set candidates.

For example, some of work (e.g., [40]) try using the existing person re-identification dataset for person search with attribute. In addition, pedestrian data set with rich annotation on attributes has been recently released [41], so the authors may use the data set too.

[40] Layne R., Hospedales T.M., Gong S. (2012) Towards Person Identification and Re-identification with Attributes. In: Fusiello A., Murino V., Cucchiara R. (eds) Computer Vision - ECCV 2012. Workshops and Demonstrations. ECCV 2012. Lecture Notes in Computer Science, vol 7583. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33863-2_40

[41] @article{li2019richly,
title={A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios},
author={Li, Dangwei and Zhang, Zhang and Chen, Xiaotang and Huang, Kaiqi},
journal={IEEE transactions on image processing},
volume={28},
number={4},

pages={1575--1590},
year={2019},
publisher={IEEE}
}

4. There are many editorial errors.

The reviewer found a number of editorial errors even from the first page as below. The authors should have done extensive proofreading before submission.

- Paper title: The authors miss a prefix (i.e., ALGORITHMS: or APPLICATIONS:), which is instructed to be added to the paper title in the author instruction document.

- l. 113: cloth-type -> clothes-type

- l. 121: Current, trend -> current trend

- l. 131: ``0.729 %w IoU > 0.4'' This part does not make sense without elaboration.

- l. 143: cloth color -> clothes color

- l. 160: might not be suitable -> might be unsuitable

- l. 183: Common nouns should not be capitalized (e.g., Person Search -> person search, Person Re-identification -> person re-identification).

- l. 184: image query -> an image query

- l. 186: than -> then

- l. 186: retrieved -> is retrieved

- l. 205: view and distance invariant -> view- and distance-invariant

- l. 206: predicting color -> predicted color

- A performance criterion %w IoU is not elaborated.

- Figure 4: Legends and labels for axes in each graph are too small to read.

5. FINAL RECOMMENDATION Please use "Borderline" sparingly.

Weak Reject

6. JUSTIFICATION Justify your final recommendation based on the strengths and weaknesses. Please be considerate to the authors and provide constructive feedback.

Since this paper suffers from a number of concerns rather than enjoys strengths, the overall rating inclines to rejection.